

GENERATING STATISTICS ON TEXT PATTERN MATCHING

PREDICATES FOR ACCESS PLANNING

Abstract of the Disclosure

Statistics for a pattern matching predicate are generated using stored character statistics. A first structure stores, for each of a plurality of character positions, frequently occurring characters in that character position, and a count of the number of occurrences of that character. A second structure stores frequently occurring characters that are subsequent to the frequently occurring characters stored in the first structure, and a probability of occurrence of each frequently occurring subsequent character. To form an estimate of the number of tuples matching a pattern matching predicate, statistics are retrieved for the matching characters in each matching position in the predicate, and then combined to produce the estimate. In the event a statistic is not stored for a desired character, the available statistics are used to make an estimate by accumulating statistics for other characters, and then calculating average frequency of occurrence of characters that do not have stored statistics.